Some models in evolutionary biology

Guillot, Gilles

Publication date:
2012

Document Version
Publisher's PDF, also known as Version of record

Some models in evolutionary biology

Gilles Guillot

Department of Informatics and Mathematical Modelling
Technical University of Denmark
Copenhagen, Denmark
http://www2.imm.dtu.dk/~gigu/

August 2012

Outline

1. Spatial clustering model(s) for the analysis of population genetic structure
2. From clustering models to hybrid zones
3. A "unifying model" for the analysis of genetic, phenotypic and spatial data
4. "Causal modelling" of genetic or phenotypic differentiation: dismantling the Mantel tests
5. Detection of correlation between genotypes and environmental variables

Spatial clustering models

What do clustering models do?

- Data:
  - co-dominant or dominant markers at neutral unlinked loci
  - individual genotypes (or allele counts over some sampling units)
  - extra information thought to be useful (e.g. sampling locations)
- Output:
  - NO ADMIXTURE MODEL / MIXTURE MODEL
    origin of individuals (or population) in one of K units
  - implicitly an estimate of K
  - ADMIXTURE MODEL
    origin of each allele (admixture model)
  - average proportion of each individual's genome in one of the K clusters

What are these clusters?

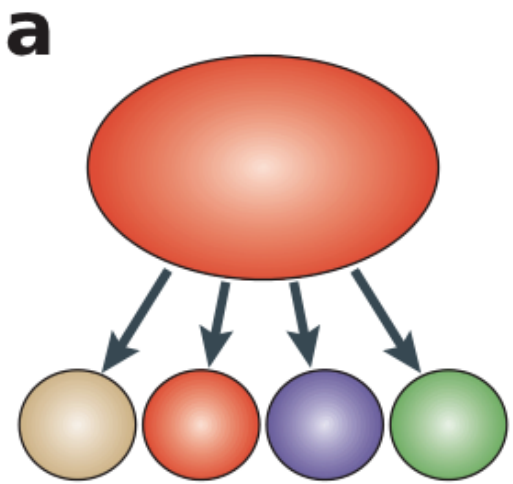
In the pioneering work of [Pritchard et al., 2000]:

- Each cluster is assumed to be at Hardy-Weinberg equilibrium
- Allele frequencies vary across cluster

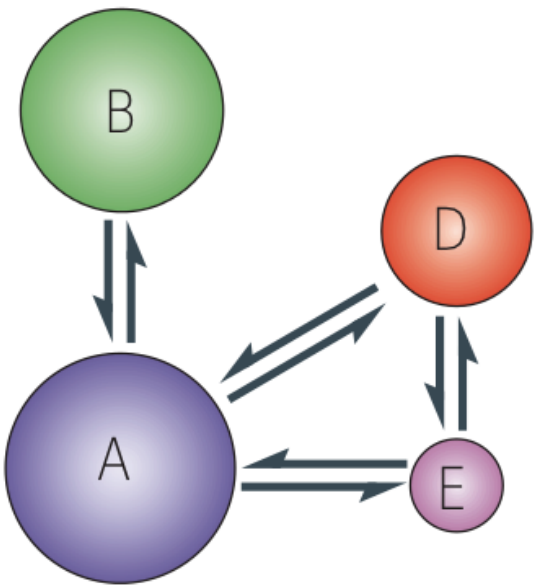
When does this make sense?

- Panmixia
- Limited gene flow between clusters

"Biological scenario" underlying clustering models



Split of an ancestral population



Islands with limited gene flow

Figure reprinted from [Hey and Machado, 2003]

What clustering models do *not* do

- Use an explicit biological model? NO
- Infer an evolutionnary scenario? NO
- Discriminate bewteen the above scenario? NO

The simplest clustering problem on Earth

- n individuals genotyped at a single bi-allelic locus
- Number of clusters K known to be 2
- Find $c_1, ..., c_n$ $c_i = 1$ or 2

Imagine allele frequencies are known

- f_1 minor allele frequency in cluster 1, f_2 m.a.f. in cluster 2
- For a given clustering,
  - n_1^{aa} homozygous aa in cluster 1
  - n_1^{aA} heterozygous aA
  - n_1^{AA} homozygous AA
  - n_2^{aa} homozygous aa in cluster 2
  - n_2^{aA} heterozygous aA
  - n_2^{AA} homozygous AA

For this given clustering, the probability of the dataset is:

$$\text{Prob}(\text{Data} \mid f_1, f_2, c_1, ..., c_n) = (f_1^2)^{n_1^{aa}} \times (2f_1(1-f_1))^{n_1^{aA}} \times ((1-f_1)^2)^{n_1^{AA}} \times (f_2^2)^{n_2^{aa}} \times (2f_2(2-f_2))^{n_2^{aA}} \times ((2-f_2)^2)^{n_2^{AA}}$$

So, if allele fequencies in the two clusters were known...

Estimating clusters could be done by finding the clustering that maximizes

$$\begin{aligned}\text{Prob}(\text{Data}|f_1, f_2, c_1, ..., c_n)) &= (f_1^2)^{n_1^{aa}} \times (2f_1(1-f_1))^{n_1^{aA}} \times ((1-f_1)^2)^{n_1^{AA}} \\ &\times (f_2^2)^{n_2^{aa}} \times (2f_2(2-f_2))^{n_2^{aA}} \times ((2-f_2)^2)^{n_2^{AA}}\end{aligned}$$

Actually, allele fequencies are not known

Cluster memberships and allele frequencies have to be estimated at the same time

How?

Using the Bayes formula:

$$\text{Prob}(f, c|\text{Data}) = \text{Prob}(\text{Data}|f, c)\text{Prob}(f, c)/\text{Prob}(\textit{Data})$$

Actually, allele fequencies are not known

Cluster memberships and allele frequencies have to be estimated at the same time

How?

Using the Bayes formula:

$$\text{Prob}(f, c|\text{Data}) \propto \text{Prob}(\text{Data}|f, c)\text{Prob}(f, c)/\text{Prob}(\textit{Data})$$

Actually, allele fequencies are not known

Cluster memberships and allele frequencies have to be estimated at the same time

How?

Using the Bayes formula:

$$\text{Prob}(f, c|\text{Data}) \propto \text{Prob}(\text{Data}|f, c)\text{Prob}(f, c)$$

Generically:

$$\textit{Posterior} \propto \textit{Likelihood} \times \textit{Prior}$$

Two terms

The likelihood $P(\text{Data}|f, c)$

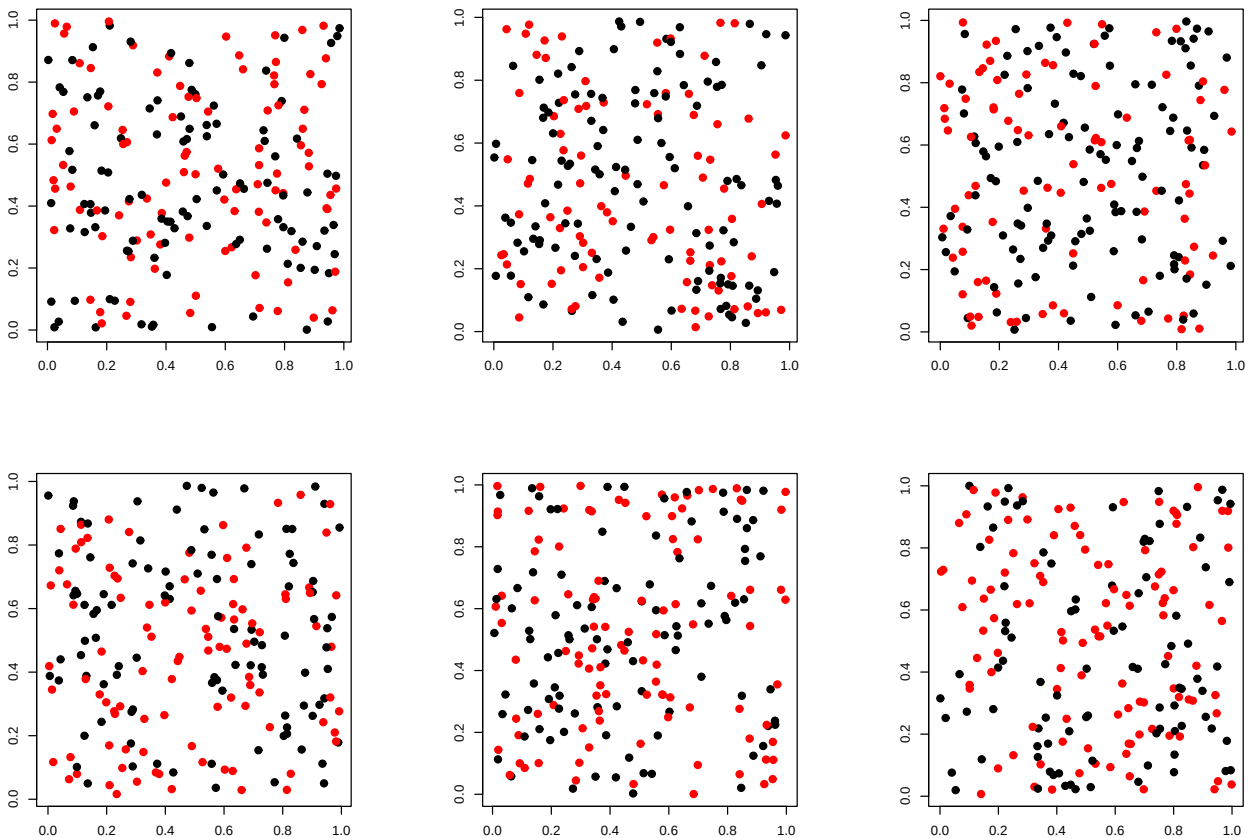
- Derives from HWE
- Results from a choice, but compulsory under HWE
- Same for mixture and admixture models
- Common to *all* clustering models: ADMIXTURE, BAPS, GENELAND, SABER, STRUCTURAMA, STRUCTURE, TESS...
- Related likelihood in INSTRUCT and GENECLUST

The prior distribution $P(f, c) = P(f)P(c)$

- Something we choose
- Should reflect our knowledge about c and f
- Main source of differences between clustering programs

Making clustering models spatial

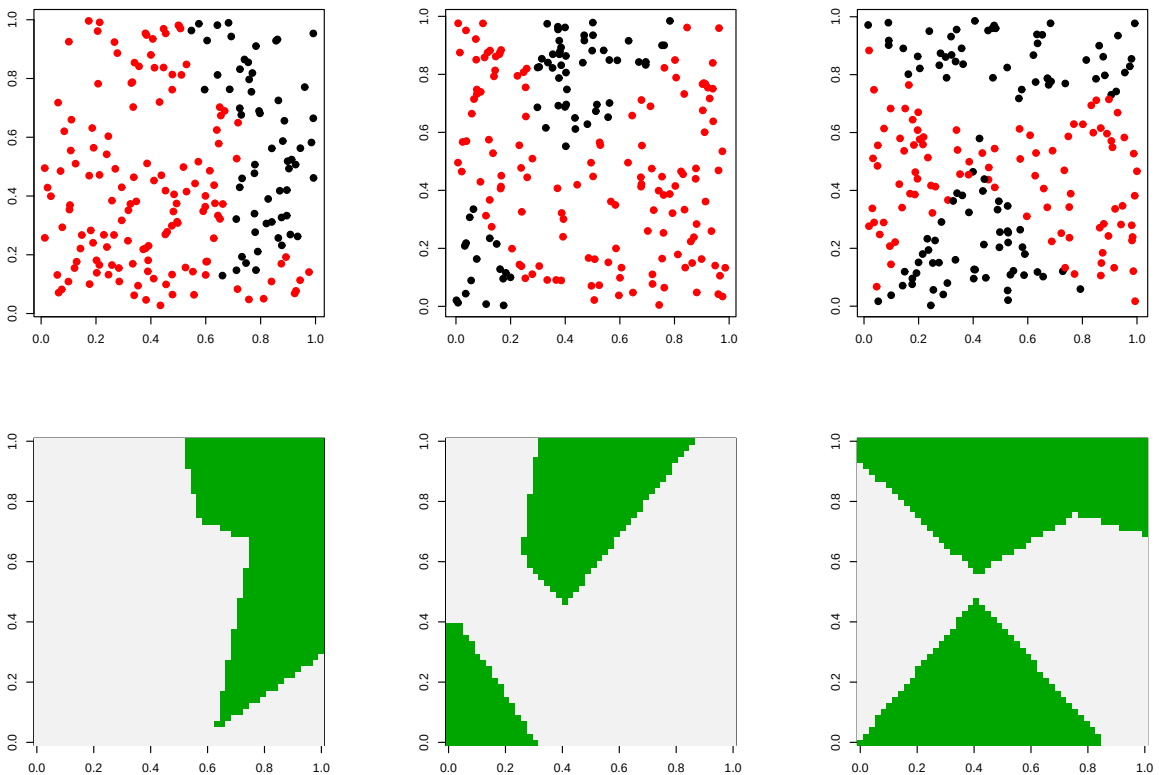
Spatial pattern of two clusters in six (simulated) samples



● cluster 1 / ● cluster 2

- Previous simulations obtained by assuming that all clusterings are equally likely
- Do not make use of spatial information whatsoever
- Exactly what is assumed in *all* non-spatial clustering models

Cluster memberships as coloured Poisson-Voronoi tessellation



Parameterization

Population membership is modelled through some auxiliary variables: number , locations and colours of polygons.

- Number of polygons: $m \sim \text{Poisson}(\lambda)$
- "Centre" of i-th polygons: $u_i \sim \text{Uniform}(D)$, u_i i.i.d
- Cluster memberships (colours): $\overset{i.i.d}{\sim} (\{1, ..., K\})$

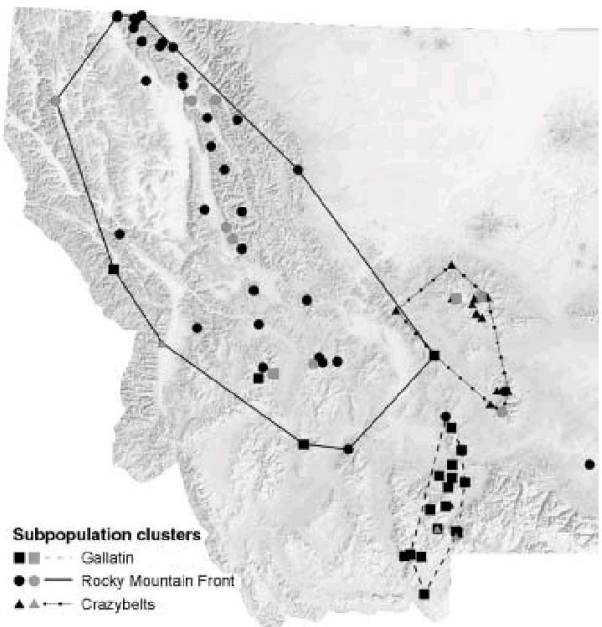
Benefit of using this model (versus non-spatial prior)

- Often better at exploring parameter space
  - more efficient in case of low genetic differentiations between clusters
  - faster convergence
  - avoid detection of spurious clusters
  - allows us to obtain a geographical map of clusters
- Some caveats
  - Prior not influential if L (nb. of loci) large
  - Poisson-Voronoi prior sometime not adapted
  - Geostatistical proverb:"Put a map on the table, and people will believe it is the truth".

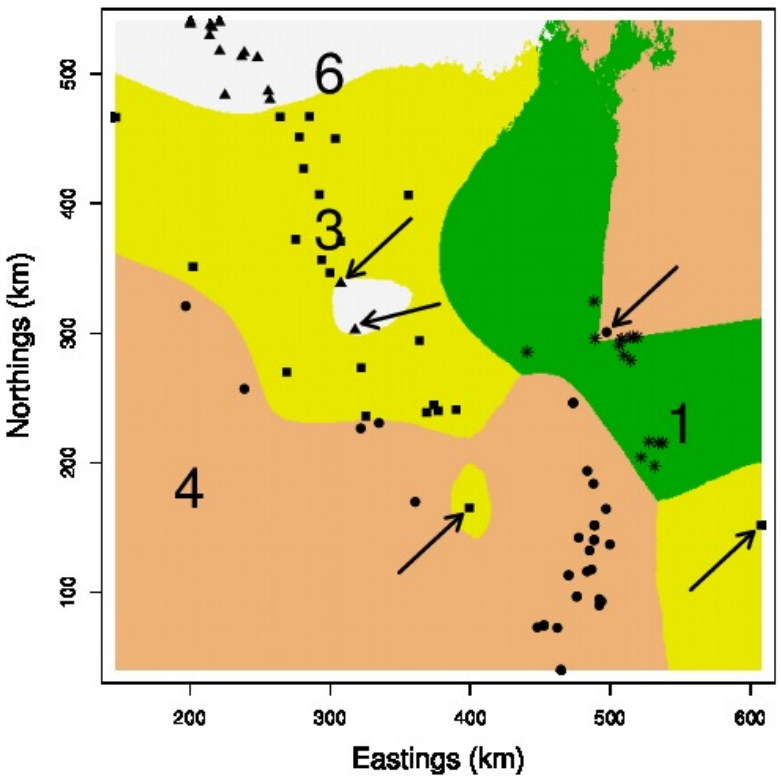
Use of spatial model above best suited for small datasets with individuals sampled regularly in space.

Montana wolverine *Gulo gulo* in North-Western US

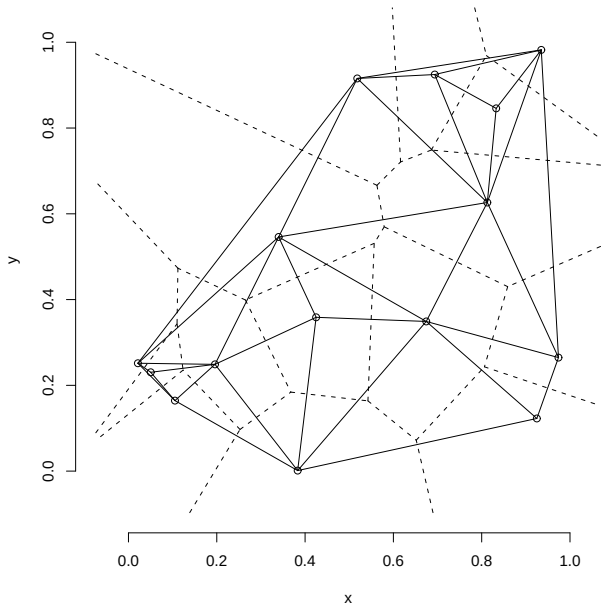
Map of inferred clusters with STRUCTURE [Cegelski et al. , 2003]



Inferred map with Geneland

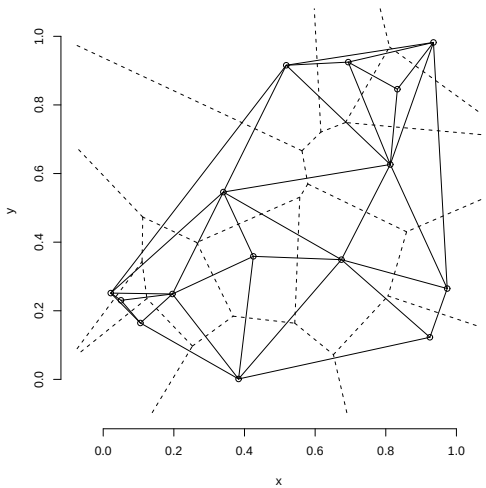


Spatial models based on a graph (or network)



- Delaunay graph
- Markov Random Field model

Pros and cons



- Graph induced by geographical distances often arbitrary
- Makes sense if the habitat has a genuine graph structure

Prior model for allele frequencies

Uncorrelated frequency model

- f_{klj} frequency of allele j at locus l in population k
- The $f_{kl.} = (f_{kl1}, ..., f_{klJ})$ are assumed **independent across populations** (and loci)
- $f_{kl.}$ Dirichlet distributed
- Unrealistic since often $f_{klj} \approx f_{k'lj}$

- Heuristic interpretation:
  - Fictional ancestral population P_A with allele frequencies f_{Alj}
  - Assume that present-time populations derives from the split of P_A
  - Assume random drift of populations parametrised by $d_k \in [0, 1]$

Alternative: the correlated frequency model

- f_{lj}^A set of $D(1, ..., 1)$ allele frequencies
- $(d_1, ..., d_K)$ vector of i.i.d uniform drift parameters
- $f_{kl.}|f^A, d \sim \text{Dirichlet}\left(f_{l1}^A(1-d_k)/d_k, ..., f_{lJ_l}^A(1-d_k)/d_k\right)$.
- Across-population correlation: $Cor(f_{klj}, f_{k'lj}) = \frac{1}{1+E[d_k]\frac{E[f_{lj}^A]-E[(f_{lj}^A)^2]}{E[(f_{lj}^A)^2]-E[f_{lj}^A]^2}}$

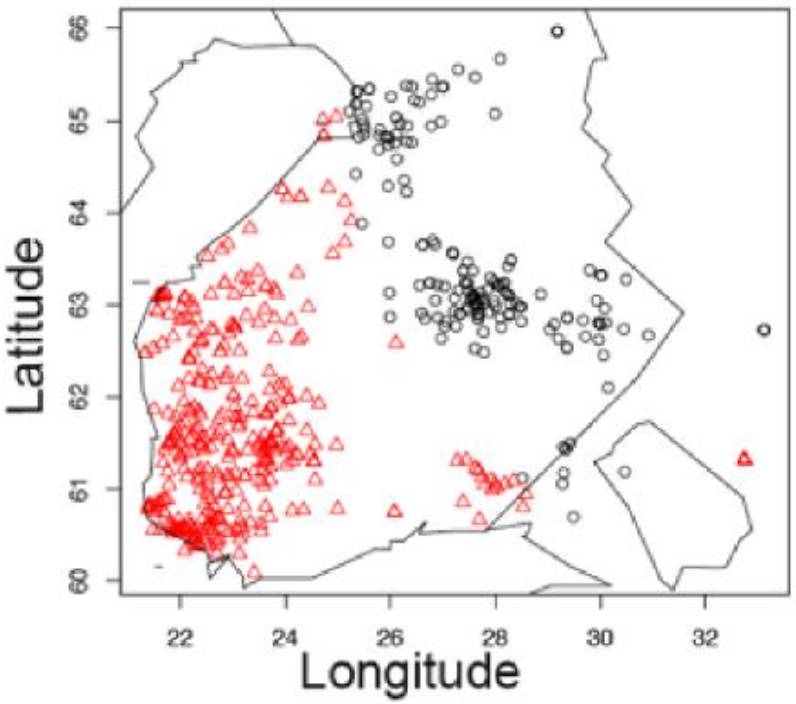
Integrated likelihood $\pi(y|c) = \int \pi(y|c, f)\pi(f)df$ depends on $\pi(f)$

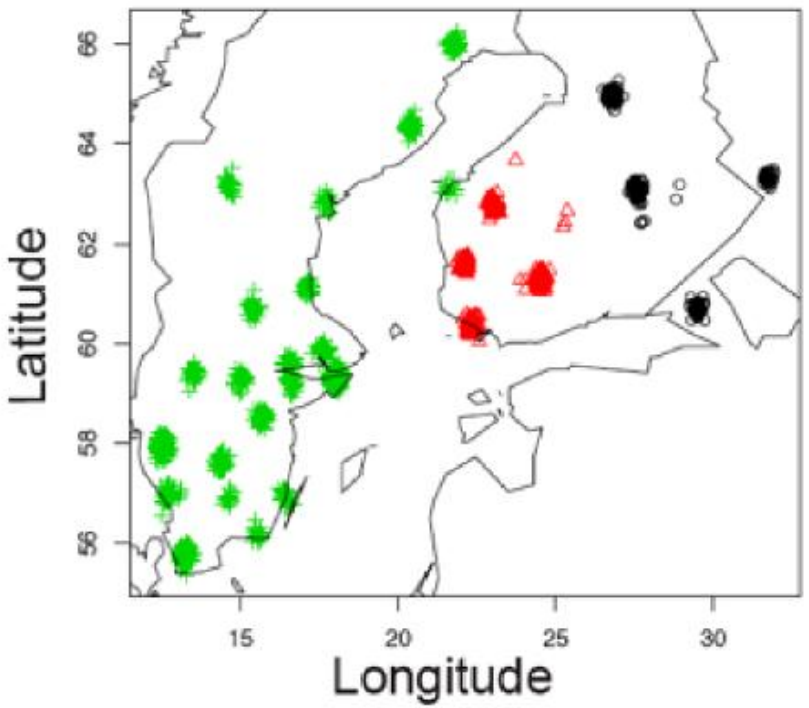
Pros and cons

- Presumably more powerful in case of low differentiation
- Does not rely on an a biological scenario
- More parameters to estimate (computing time)
- Prone to numerical instabilities

Example of increased power of the correlated frequency model: human genetic data in Finland and Sweden

34 autosomal SNPs, 2701 individuals + 40 microsat. on 465 individuals F_{ST} ranging 0.01-0.001 [Hannelius et al., 2008]





Summary of the various modelling assumptions in spatial clustering models:

Assumptions on genetic features within populations

- Hardy-Weinberg equilibrium
- Linkage equilibrium
- Possible correlation between allele frequencies

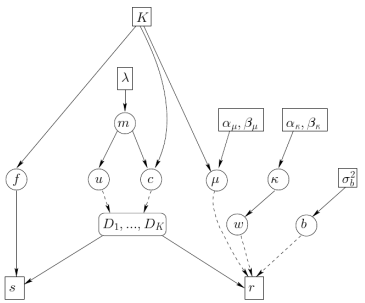
Assumptions on the spatial population spread

- no genetics, but...
- ... must be consistent with genetic assumptions
- i.i.d prior non spatial model
- polygonal population areas
- graph based models

Inference algorithm

Simulation based method

- MCMC

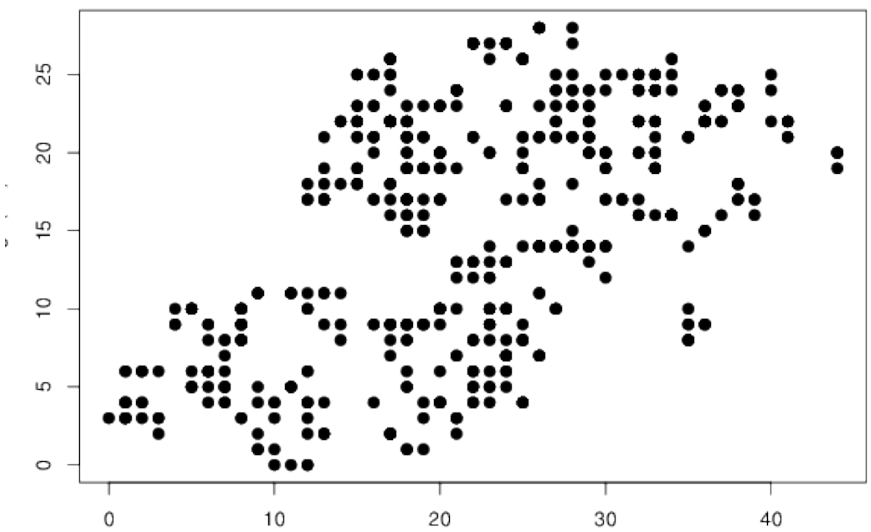


Dependence, multivariate

- Transdimensional algorithms
- Post-processing issues of MCMC outputs
  label switching issue

Accounting for uncertainty about spatial coordinates

Roe deer *Capreolus capreolus* in the South-West of France



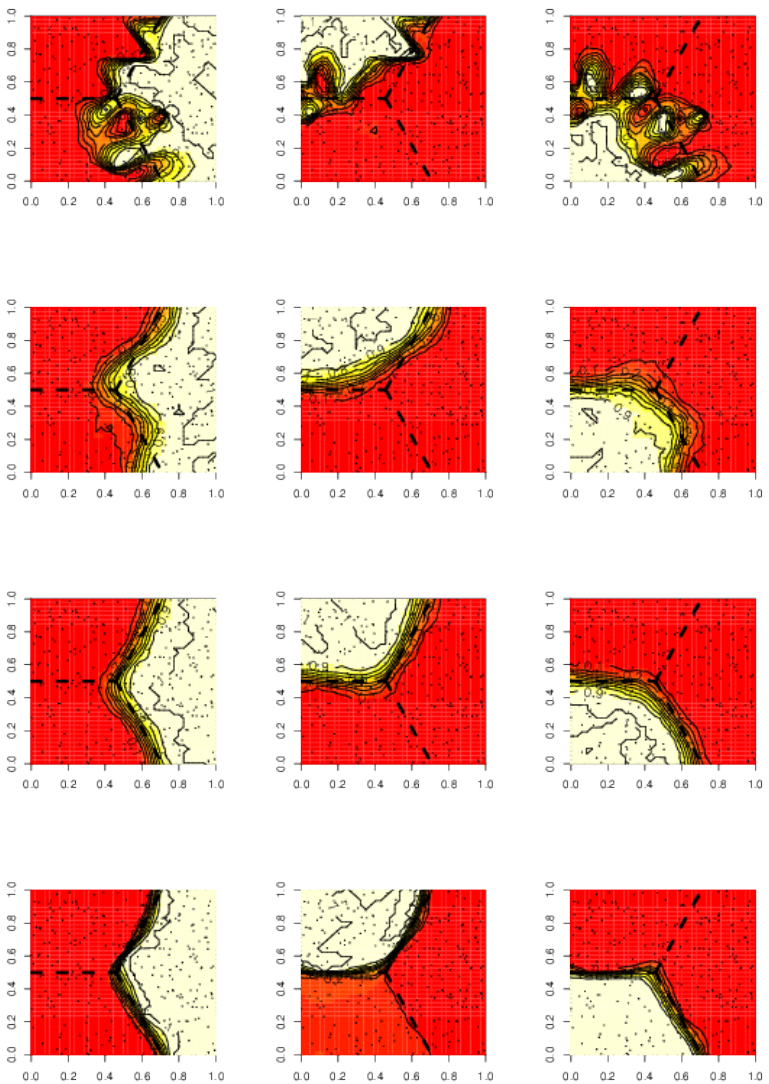
Uncertainty about spatial coordinates

Recorded coordinates s_i may be
- blurred by noise
- coarsely recorded
- uncertain due to animal movements
- not fully meaningful for animals with large home range
- problematic in case of individuals sharing the same spatial location

Introduce some "true" (unobserved) coordinates t_i
They are related to the observed coordinates s_i by

$$s_i = t_i + \varepsilon_i$$

where ε_i is an i.i.d additive noise chosen in a *suitable* parametric distribution.



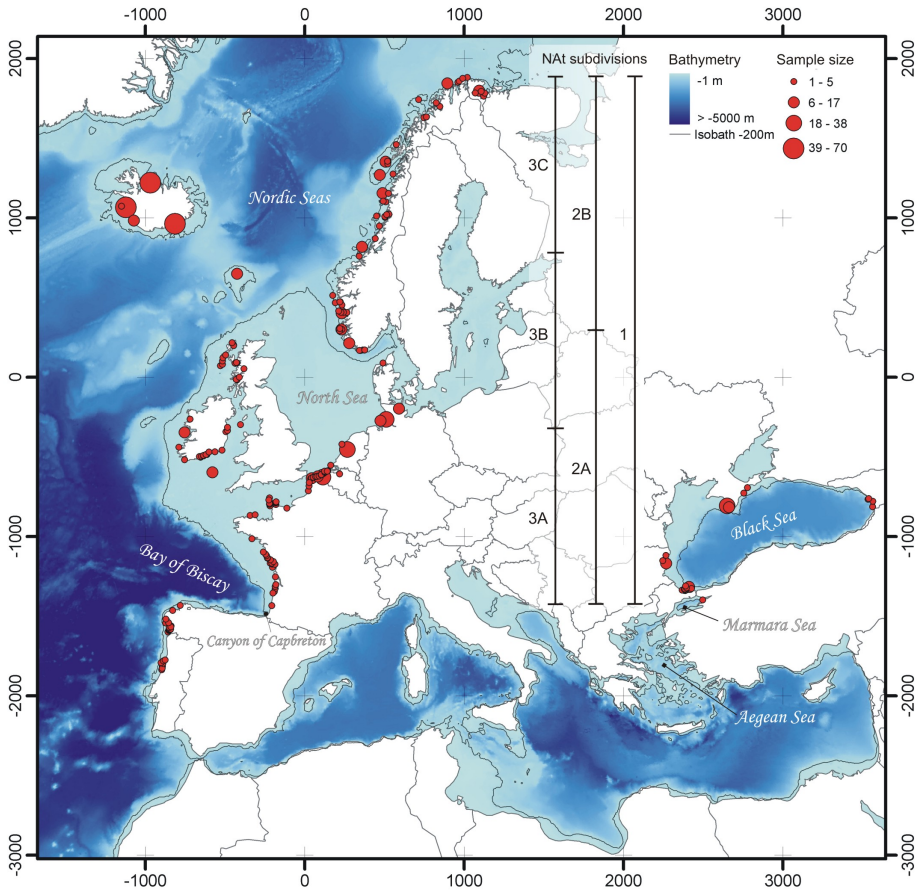
Dominant markers

- Two alleles a, A

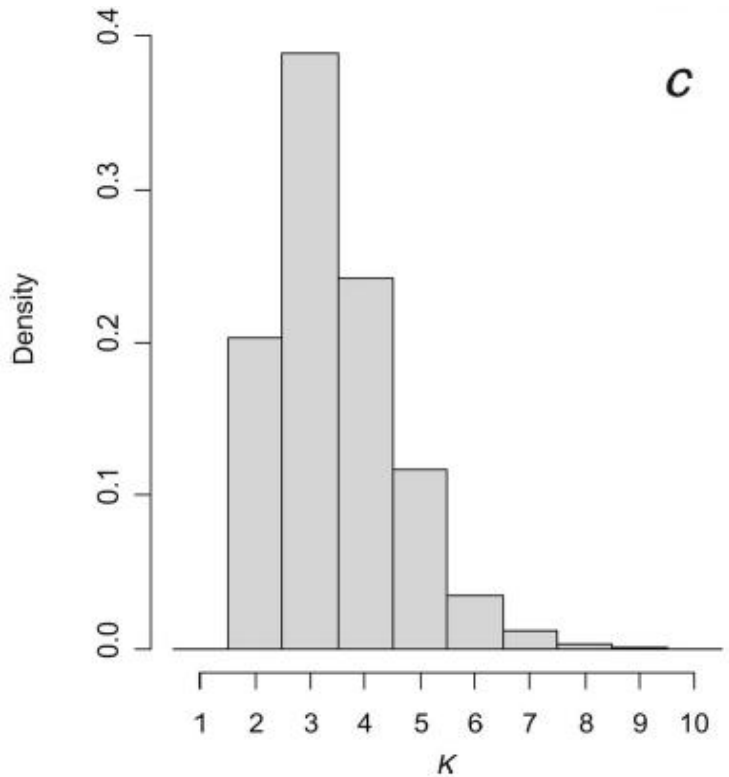
Genotype	Data
A, A	Presence
A, a	Presence
a, a	Absence

  (A, A) and (A, a) can not be distinguished
- Uncertainty can be handled in statistical computations
- Obvious loss of accuracy, but how much?
- Magic number: 1.69

The North-Atlantic harbour porpoise data [Fontaine et al., 2007]

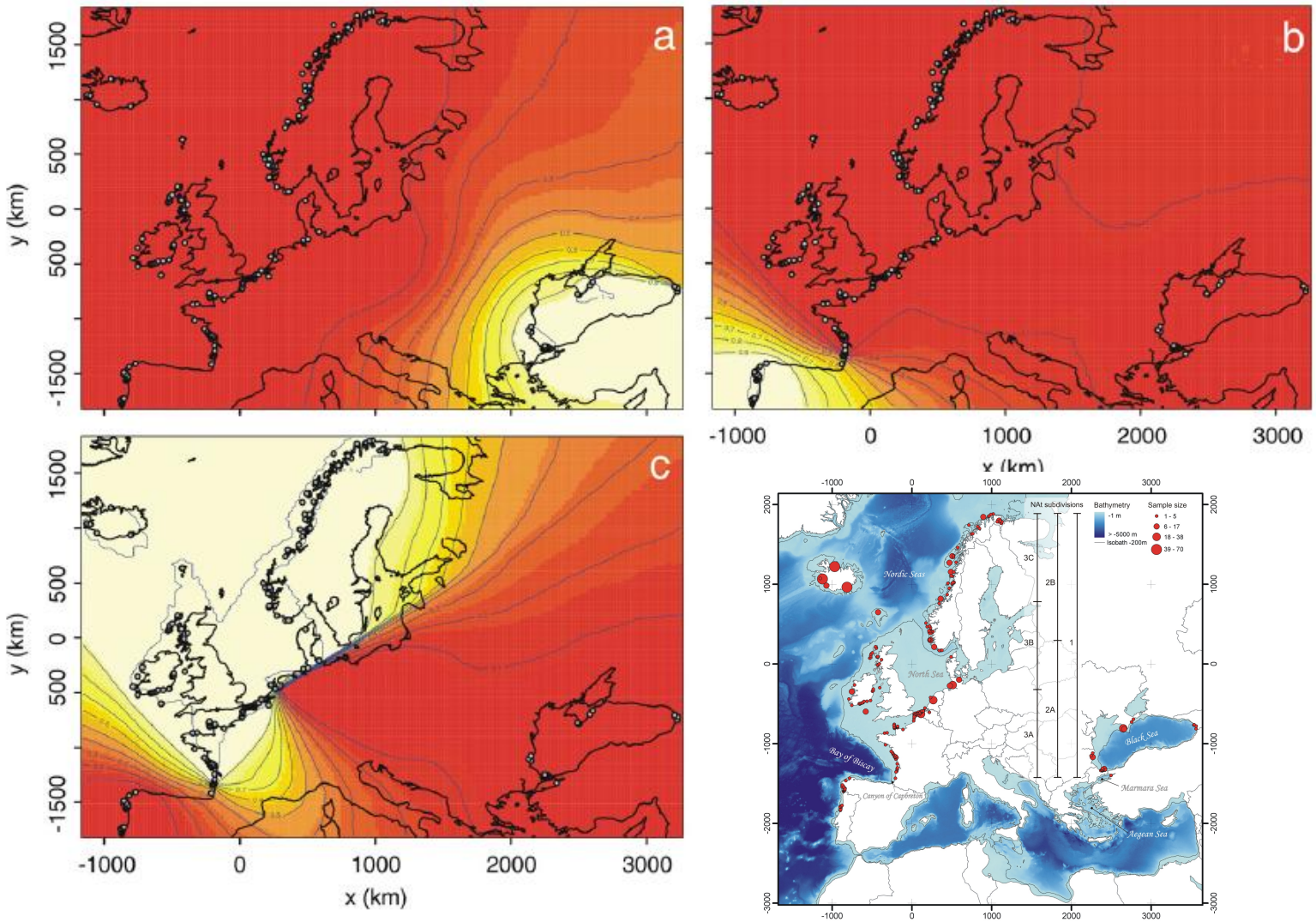


Inferred number of clusters K for the harbour porpoise data

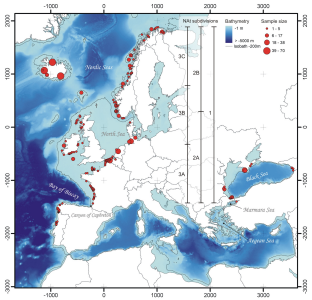


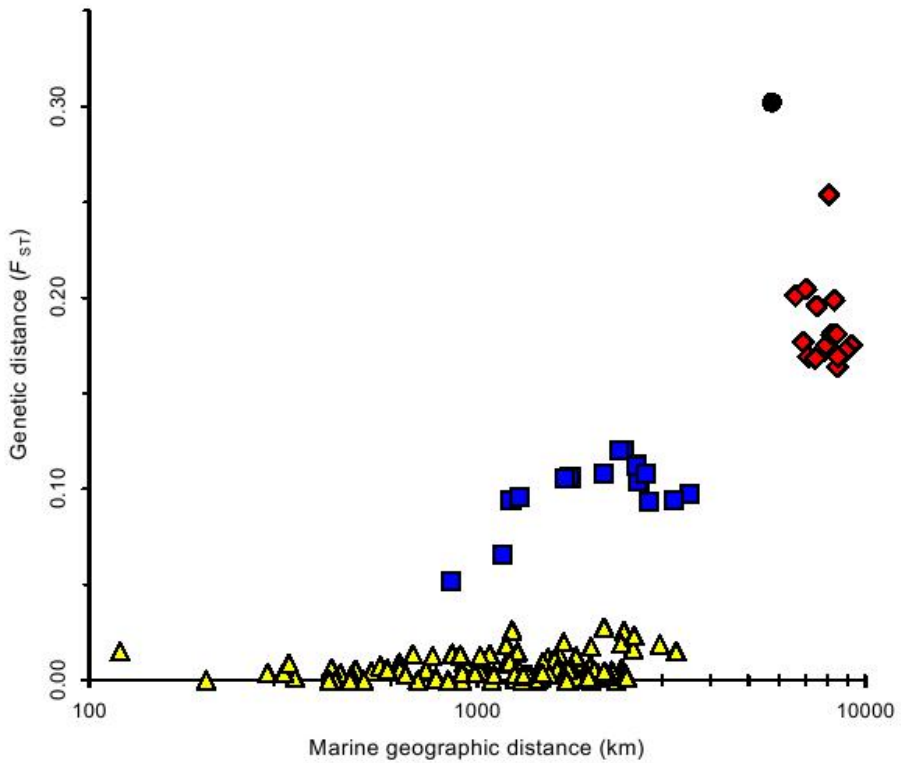
Posterior distribution of number of clusters K for the harbour porpoise data. The mode is $\hat{K} = 3$.

Inferred structure for porpoise data



Evidence of clines and clusters [Fontaine et al., 2007]





From clustering models to hybrid zones

Modelling admixture

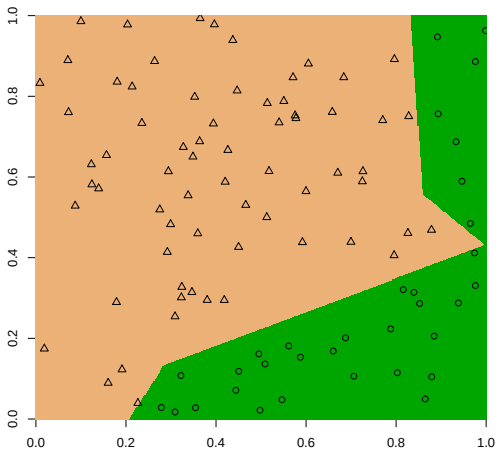
- Individuals with admixed ancestries
- No attempt to recover the true evolutionary scenario
- Object of inference: proportion of individuals genomes in the various clusters
- Model pioneered by Pritchard et al. 2000 and Falush et al. 2003 works great.
- MCMC-free (EM algorithm) approach by Alexander et al., 2009 (ADMIXTURE)

Admixture coefficients

Definition of admixture coefficients

- $q_{i.} = (q_{ik})_{k=1,...,K}$ proportion of alleles carried by i with origin in cluster k
- Have to sum-up to one: $\sum_k q_{ik} = 1$
- Calls for Dirichlet distribution $\mathcal{D}(\alpha_{i1}, ..., \alpha_{iK})$

Dirichlet admixture coefficients with spatially varying hyper-parameters

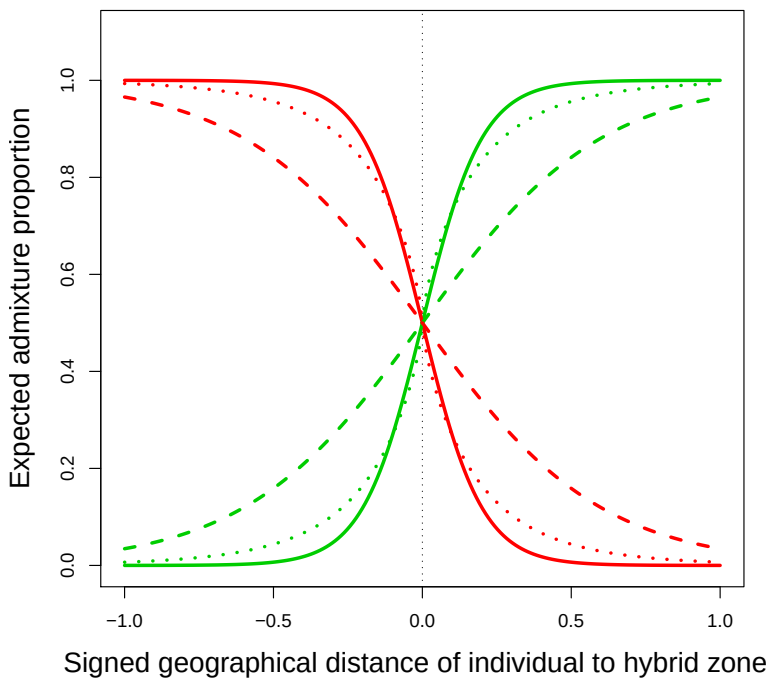


Main assumption

- $q_{i.} = (q_{ik})_{k=1,...,K} \sim \mathcal{D}(\alpha_{i1}, ..., \alpha_{iK})$
- $\alpha_{ik} = a \exp(-d_{ik}/b)$ d_{ik} distance of indiv. i to cluster k

Admixture proportions relates to distance to the contact zone as:

$$E[q_{ik}] = \frac{e^{-d_{ik}/b}}{\sum_k e^{-d_{ik}/b}}$$



Interpretation of b parameter: some limiting cases

Coming back to the definition

$$\alpha_{ik} = a \exp(-d_{ik}/b)$$

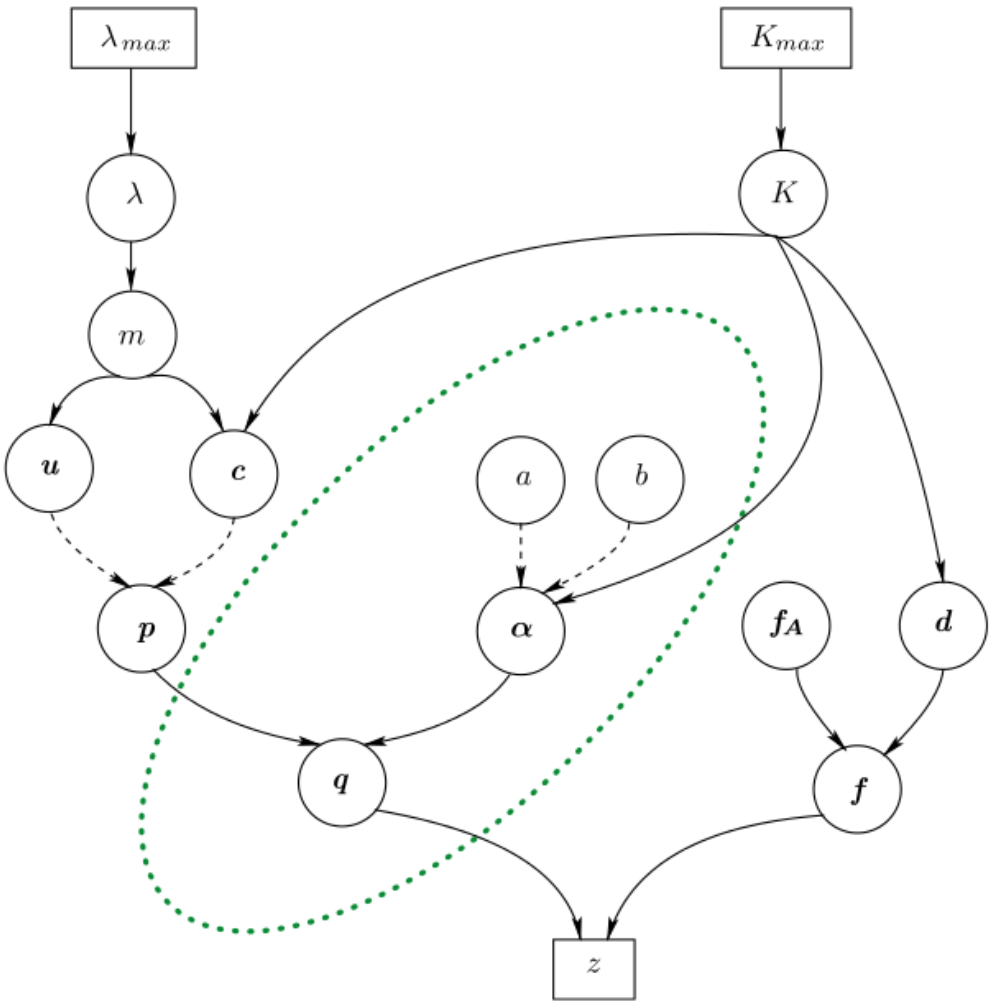
- for large b, $\alpha_{ik} \approx a$
  $(q_{ik})_{k=1,\ldots,K} \sim \mathcal{D}(a, \ldots, a)$
  no spatial structure
- for $b = 0$, $\alpha_{ik} \approx 0$ or 1
  strong spatial structure

Interpretation of a parameter

$$\alpha_{ik} = a \exp(-d_{ik}/b)$$

- $E[q_{ik}] = \dfrac{e^{-d_{ik}/b}}{\sum_k e^{-d_{ik}/b}}$
  $\Longrightarrow$ a not involved in the expected admixture coefficient
- a involved in the variance of q_{ik}: $V[q_{ik}] \propto 1/a$

Graph of model with hybrid zone

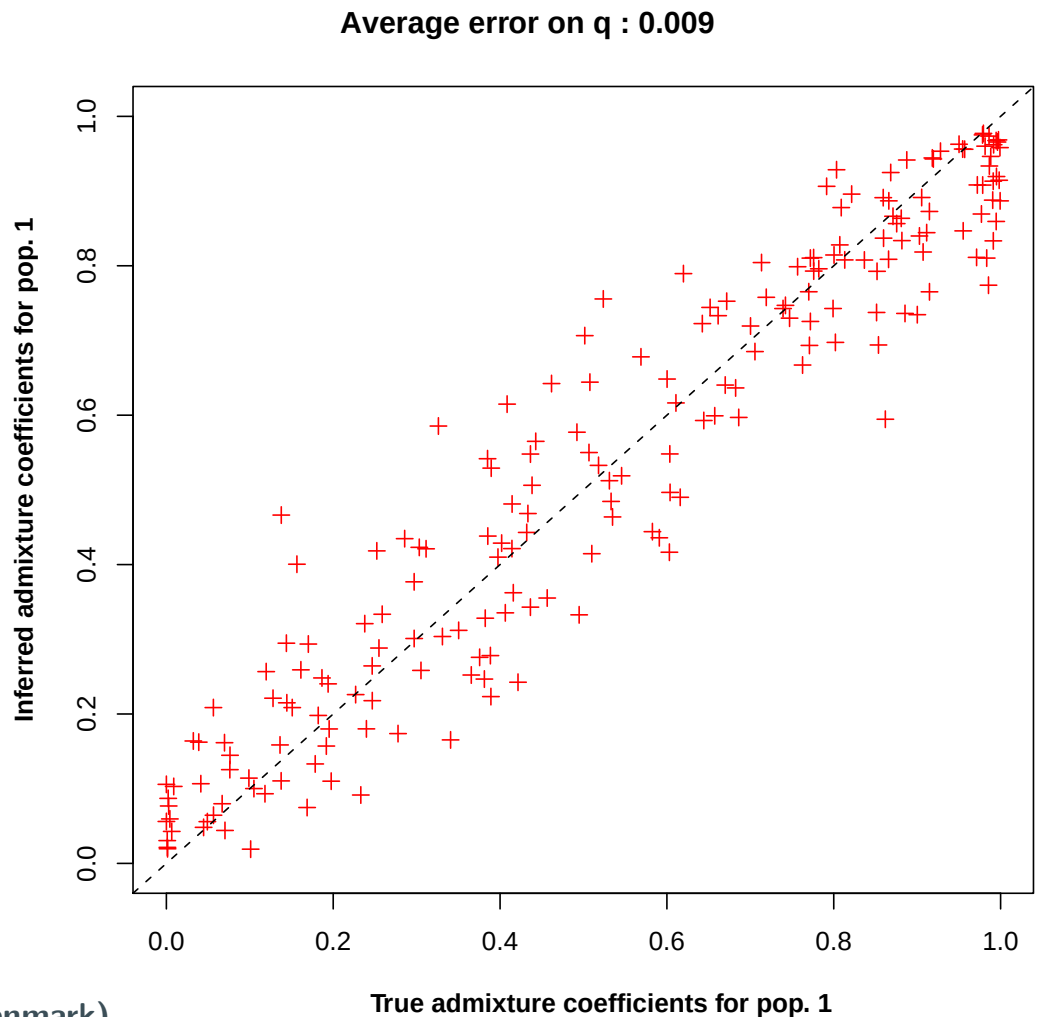


Parameter inference

- Full Bayesian inference out of reach
  - time consuming to develop
  - time consuming to run
  - presumably highly prone to usual MCMC issues
- Two-step procedure (reminiscent of [Macholan et al. 2011])
  - First MCMC run under no-admixture model
    provides information about number of clusters, locations of contact zones and allele frequencies
  - Second run to infer hybrid zone parameters b (and possibly a)

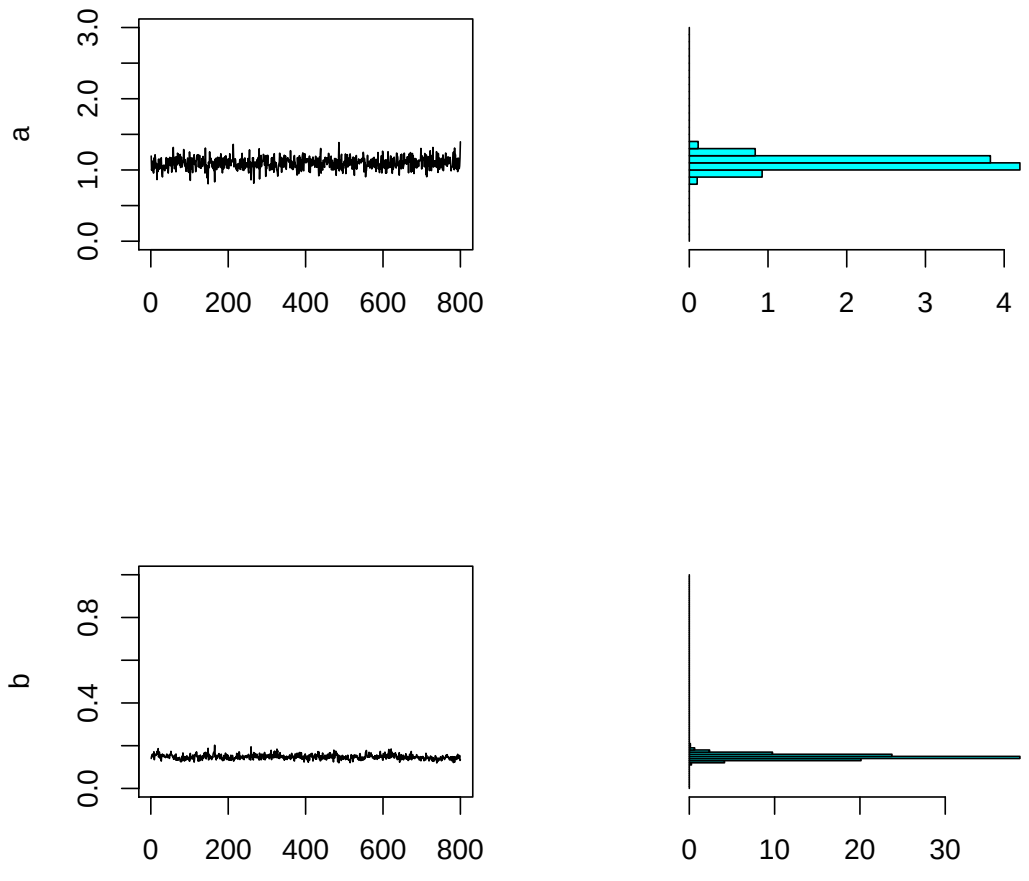
Example of parameter inference output

Estimated versus true (simulated) admixture coefficients



Example of parameter inference output cont'

MCMC trace and posteriod distribution of parameters *a* and *b*



Clustering morphometric data

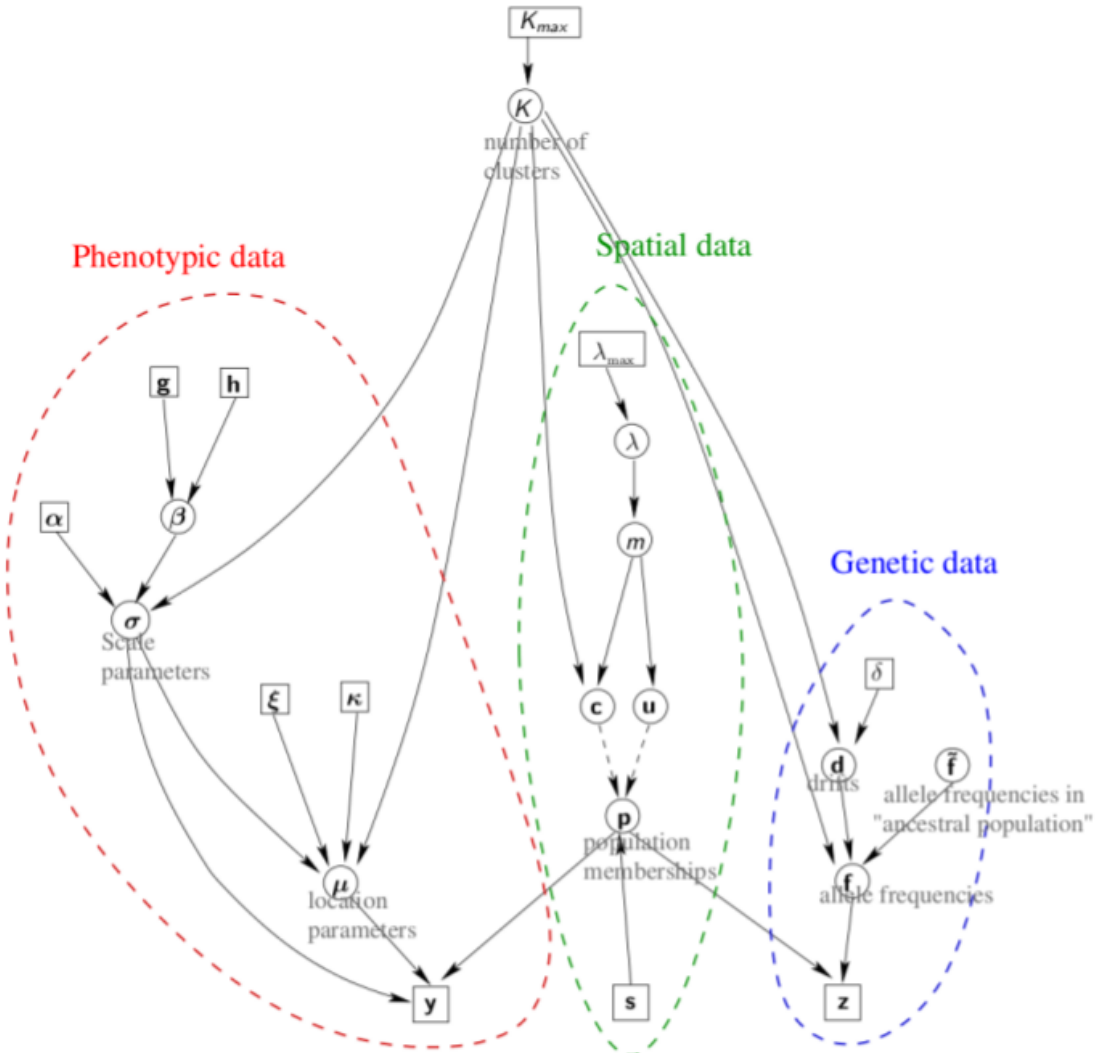
Clustering model for morphometric data

- Work prompted by Geneland users dealing with morphometric data
- Clustering models for quantitative variables predate work in Pop. Gen. but application wiht morphometric data are scarce

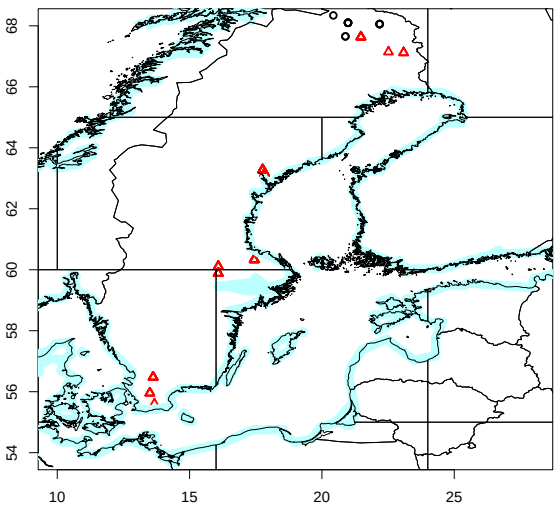
Main goals of proposed model

- Cast inference framework for phenotypic data similar to that for genetic data
- Explore options for joint analysis

Graph of full model

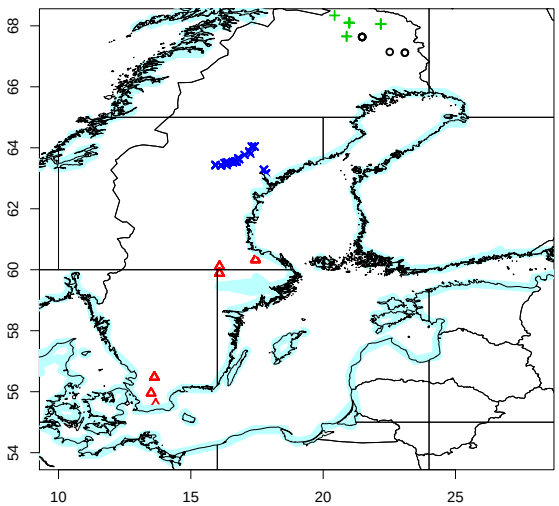


Example of application: the *Myodes* data in Sweden [Guillot et al., 2012]



Phenotypic & Spatial

$\hat{K} = 2$



Genetic & Spatial

$\hat{K} = 4$

An important caveats on the model for morphometric data

Model for genetic data not changed: still not tailored for non recombining data

Dismantling the Mantel tests

The (simple) Mantel test

Mantel N., **The detection of disease clustering and a generalized regression approach**, *Cancer Research*, 27, 209-220, 1967.

- Goal: *"identifying subtle time-space clustering of disease, as may be occurring in leukemia"*
- Data: $(x_i, y_i)_{i=1,\dots,n}$ observations of a space-time point process
- Idea:
  - transform data so as to get two univariate variables
  - compute correlation of transformed data
  - assess significance of correlation by some permutation method

The simple Mantel test: detailed algorithm

- Compute $D^x = (|x_i - x_j|)_{i,j}$ and $D^y = (|y_i - y_j|)_{i,j}$
- Compute the empirical correlation r between D^x and D^y
- For `iter =1,N`
  - draw a random permutation τ of $1, ..., n$
  - compute $D^x_\tau = (|x_{\tau(i)} - x_{\tau(j)}|)_{i,j}$
  - compute the empirical correlation r_τ between D^x_τ and D^y
- If $|r|$ larger than some quantile estimated from the r_τ values: report that there is *"subtle time-space clustering of disease"*

The partial Mantel test

- x_i and y_i observations of p and q variables for n statistical units.
- still attempts to assess the dependence between x and y
- need to *"filter out"* or *"control for"* the effect of a third variable z (e.g. z_i spatial coordinates of obs. i)

The partial Mantel test: detailed algorithm

- Compute $D^x = (|x_i - x_j|)_{i,j}$, $D^y = (|y_i - y_j|)_{i,j}$ and $D^z = (|z_i - z_j|)_{i,j}$
- Compute residuals $\tilde{D}^x$ of linear regressions $D^x \sim D^z$
- Compute residuals $\tilde{D}^y$ of linear regressions $D^y \sim D^z$
- Compute the empirical correlation r between $\tilde{D}^x$ and $\tilde{D}^y$
- For `iter =1,N`
  - draw a random permutation τ of $1, ..., n$
  - compute $\tilde{D}^x_\tau$ as above for permuted x_i values
  - compute the empirical correlation r_τ between $\tilde{D}^x_\tau$ and $\tilde{D}^y$
- Assess significance of r by comparing to quantiles of r_τ.

Mantel put into orbit

Mantel (Cancer Res., 1967) and Sokal (Sys. Zool., 1979) claimed that

- the approach was general
- could be used to assess dependence between matrices of "distance"

Features of the method

- deals with multivariate data
- synthetize data into a single numerical value
- does not seem to rely on any distributional assumption

Posterity of Mantel's work

- Simple Mantel test: $\geq$ 5000 ISI citations
- Partial Mantel test : $\geq$ 1000 ISI citations
- Implemented in most ecology computer programs
- Countless number of articles using the Mantel tests citing other supporting references
- Routinely used in landscape genetics: x genotypes, y environmental variables, z geographical coordinates
- Practice strongly rooted:

Pr. XXX, Assoc. Editor J. of XXX:

"Referee 3 pointed out some issues with the Mantel tests but they are so widely used in lansdcape genetics that this comment can be disregarded."

Is the Mantel test a statistical test?

Formal definition involves...

- A null hypothesis
- A method to derive a p-value
- Some additional distributional assumptions

Are the Mantel tests appropriate?

A common implementation:

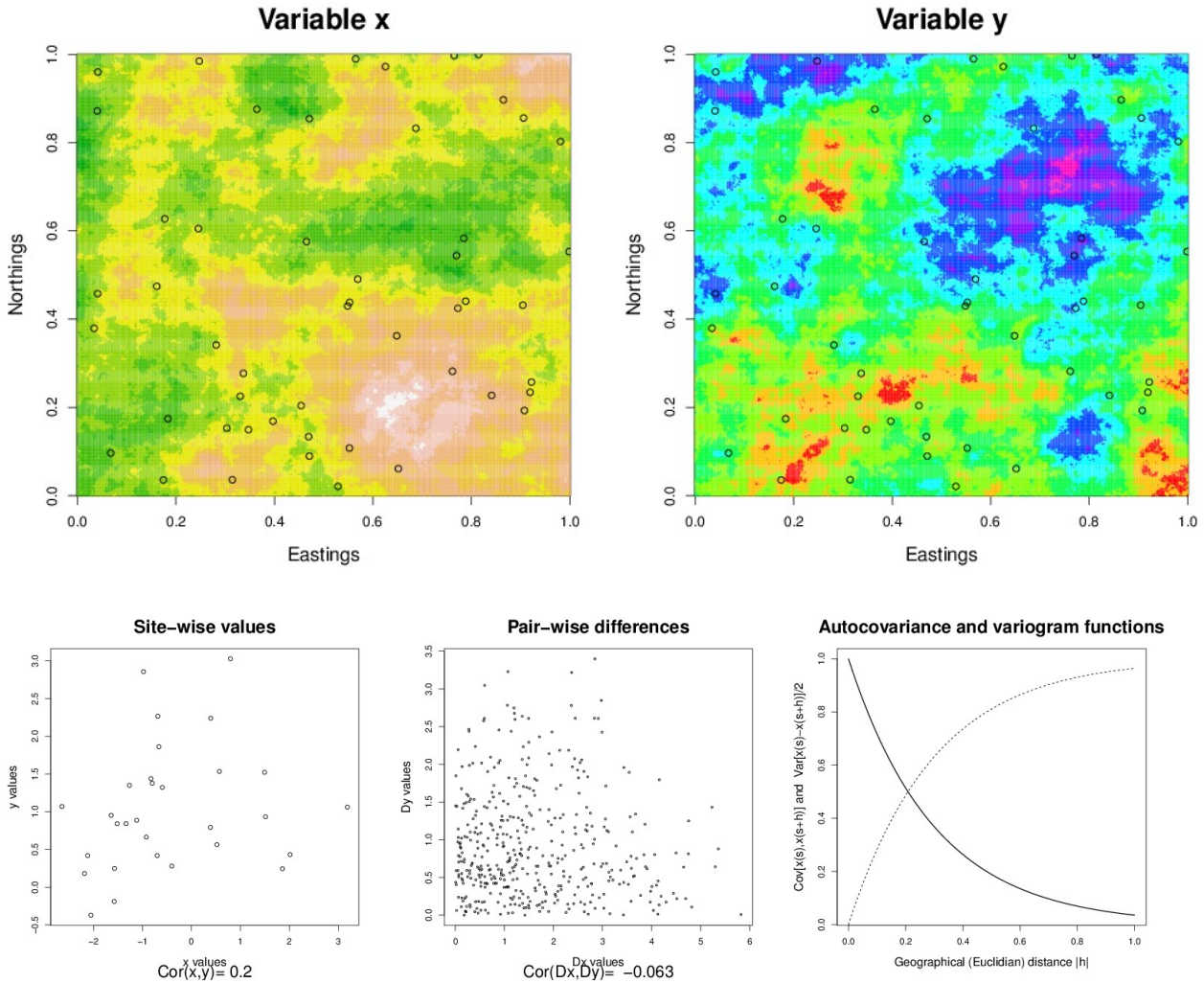
- x_i mutivariate genotype or phenotype.
  Due to population history and limited mixing in space x is spatially-autocorrelated
- y_i multivariate descriptor of landscape (elevation, temperature, vegetation cover).
  Due to bio/geo-physical laws y is spatially-autocorrelated
- Interest in testing H_0: x and y are independent

A simulation study

Simulation to mimic the situation of one phenotypic variable and one environmental variable.

- $s_1, ..., s_n$ n=50 sites in $[0, 1]^2$
- $x(s_1), ..., x(s_n)$ values of a GRF with expo. covariance
- $y(s_1), ..., y(s_n)$ values of a GRF with expo. covariance
- x and y independent
- common scale param. κ

Example of simulated data



Simulation study (cont')

- simulation above repeated for 200 realizations of x and y
- p-values for simple Mantel test
- p-value for partial Mantel test with matrix D^s entered to "control the effect of space".
- common scale param. κ vaying from 0 to 0.7
- plot of ordered p-values against quantiles of a uniform distribution
- Under H_0, the p-values should be uniformly distributed

Qq-plots of p-values obtained on simulated data

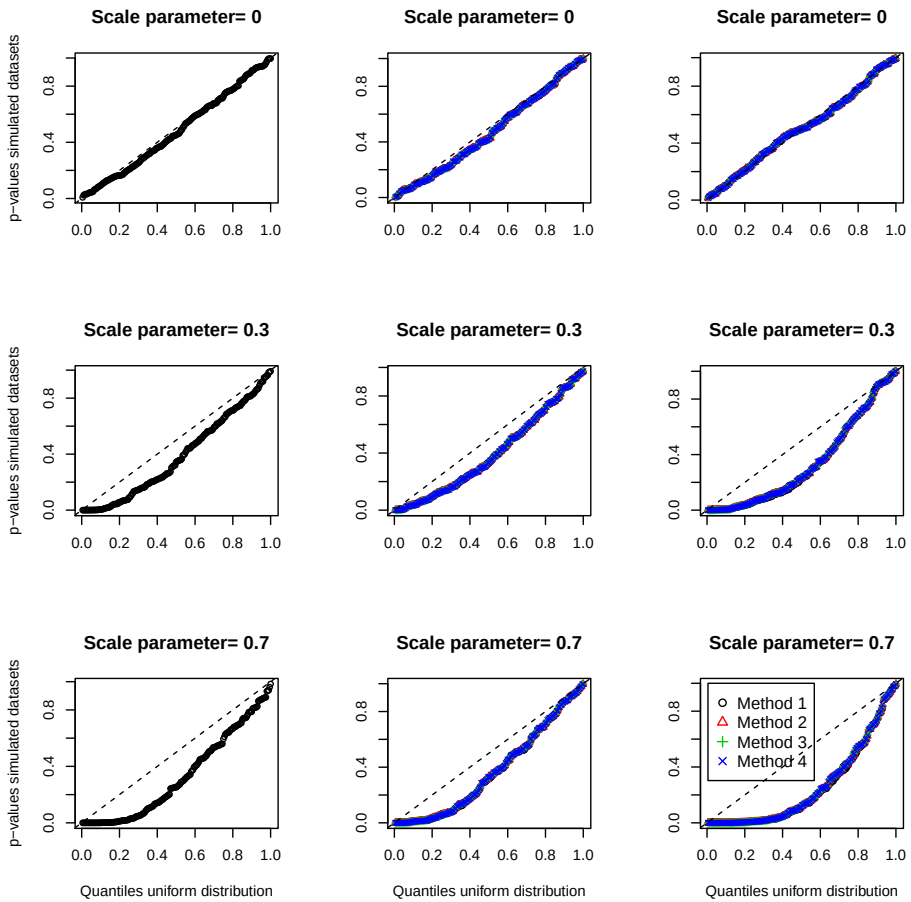


Figure: Left: simple Mantel test. Middle: partial Mantel test, no drift. Right: partial Mantel test, RFs with linear trend.

What's wrong with the Mantel tests?

Mantel tests are based on permuation of one of the data vector entries

- Permutation of x values breaks the potential dependence between x and y
- Also breaks the spatial structure of x!!

The Mantel test fallacy:

$$cor(D^x_\tau, D^y) \overset{\mathcal{L}}{\neq} cor(D^x, D^y)$$

The Mantel tests produce typically correlation coefficients of landscape descriptors with data from an island model. What we rather need is the distribution of the correlation coefficient between landscape descriptors and IBD data.

Alternative approaches

- Testing independence between two point processes [Schlather et al., 2004].
- Modified t-test to account for auto-correlation [Clifford et al., 1989, Richardson and Clifford, 1991, Dutilleul et al., 1993].
- Extension to categorical data [Cerioli, 2002]
- Restricted permutations:
  - for clumpped geostatistical data: within-population permutation
  - lattice data: shift permutation
- Testing in a GLMM framework

Conclusion

- Mantel tests are flawed in presence of structure in the data
- Conclusion extends to other form of structure (phylogenetic trees)
- A clear warning is timely
- Needs further work on the side of computer program development

Research report:
Guillot, G. and Rousset, F., On the simple and partial Mantel tests in presence of spatial auto-correlation, arXiv:1112.0651v1,(2012).

Detecting loci under selection

Background: MCMC, ABC, EM and the curse on iterative methods

- MCMC multi-purpose but time consuming to develop, run and prone to convergence issues
- ABC : approximate, high price to pay for not knowing the likelihood
- EM : tailored for missing data problems (e.g. clustering problems)
- Laplace approximation

The INLA package [Rue et al., 2009]

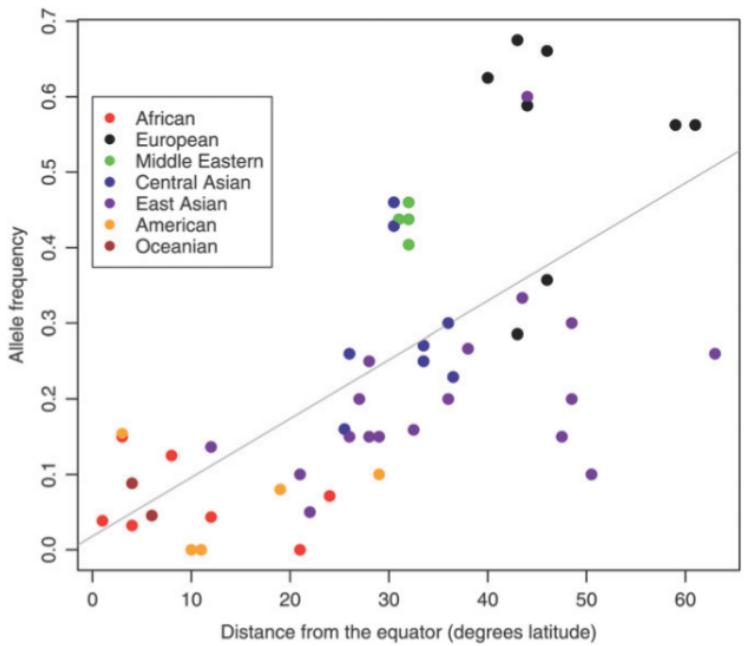
- Inference in hierarchical models
- Suitable for latent Gaussian structures
- MCMC free
- Deals with geostatistical data [Lindgren et al., 2011]

Methods of detecting selection

- Loci displaying outstanding correlation with some environmental variables.
  - provide direct functional information
  - statistically challenging

See Roger Butlin's talk for a broader perspective

Example: correlation of allele frequency with distance to the equator (reprinted from [Coop et al., 2010])



Distance from the equator for 52 human populations against sample allele frequencies (SNP AGT M235T). Points are colored according to geographic region following definitions in [Rosenberg et al., 2002].

The SAM program [Joost et al., 2007, Joost et al., 2008]

Data

- z_i allele count of pop or individual i at a bi-allelic locus
- y_i environmental variable observed at geographical location of unit i

Method: plain logistic regression

- $z_i = \text{Binom}(n_i, f_i)$
- $f_i = logit^{-1}(by_i + c)$

For short: Genotype ~ environment

- Does not account for potential spatial correlation structure due to population history
- Likely returns inaccurate p-values (cf. discussion on variable selection in [Joost et al., 2007])

The BAYENV program [Coop et al., 2010]

- Same as Joost in spirit
- But includes a random effect to account for population structure

Logistic regression

- $z_i = \text{Binom}(n_i, f_i)$
- $f_i = logit^{-1}(ax_i + by_i + c)$
- x unobserved spatially correlated random effect accounting for population history
- x Gaussian vector with inverse Wishart prior on covariance matrix
- Model above fuzzed with Nicholson's model [Nicholson et al., 2002] to account for fixed alleles

Intended to deal with genomic scans but MCMC-based inference...

The INLA-SPDE approach of detecting selection

Model reminiscent of that of Coop et al.

Spatial logistic regression with latent Gaussian structure

- $z_i = \text{Binom}(n_i, f_i)$
- $f_i = logit^{-1}(ax_i + by_i + c)$
- x unobserved spatially correlated random effect accounting for population history
- x Gaussian random field with Matérn covariance function
- INLA-SPDE approximation to treat x as a Markov random field

Summary of the INLA-SPDE method for detecting selection

Input

- Genotypes or allele counts at n sites
- Enviromental variables at "some" locations
- Spatial coordinates of measurements

Output

- Estimates of coefficients a, b and c in logistic regression
- Estimate of spatial scale of random effect
- Integrated likelihood $\int f(z|\theta)\pi(\theta)d\theta$ for various competing models (e.g. with and without spatial random effect)

INLA-SPDE vs. SAM and Bayenv

+

- INLA runs in a few seconds per locus for 100-500 individuals (minutes to hours for MCMC based methods)
- Free from MCMC convergence issues
- Returns realistic correlation estimates
- Flexible in terms of spatial sampling

Report on arxiv: Guillot, G., Detection of correlation between genotypes and environmental variables. A fast computational approach for genomewide studies. http://arxiv.org/abs/1206.0889

Conclusion

About the no-admixture model

- Well validated model and robust model for the inference of population structure
- Outperforms competing approach in landscape genetics studies [Safner et al., 2011, Blair et al., 2012]
- MCMC-based prone to convergence issues

About the hybrid zone model

- Roots some existing ideas into formal Bayesian inference
- Only approximate inference
- Model also prone to convergence issues

About the detection of loci under selection

- Work in progress with POPRES data (500k SNPs, 4000 individuals in Europe)
- Soon an R package with GUI (work with F. Santos)

Thank you!

References I

[Blair et al., 2012] Blair, C., Weigel, D., Balazik, M., Keeley, A., Walker, F., Landguth, E., Cushman, S., Murphy, M., Waits, L., and Balkenhol, N. (2012).

A simulation-based evaluation of methods for inferring linear barriers to gene flow.

Molecular Ecology Resources.

[Cerioli, 2002] Cerioli, A. (2002).

Testing mutual independence between two discrete-valued spatial processes: A correction to Pearson chi-squared.

Biometrics, 58:888–897.

[Clifford et al., 1989] Clifford, P., Richardson, S., and Hemon, D. (1989).

Assessing the significance of the correlation between two spatial processes.

Biometrics, 45(1):123–134.

References II

[Coop et al., 2010] Coop, G., Witonsky, D., Di Rienzo, A., and Pritchard, J. (2010).
Using environmental correlations to identify loci under selection.
Genetics, 185:1411–1423.

[Dutilleul et al., 1993] Dutilleul, P., Clifford, P., Richardson, S., and Hémon, D. (1993).
Modifying the t test for assessing the correlation between two spatial processes.
Biometrics, 49:305–314.

[Fontaine et al., 2007] Fontaine, M., Baird, S., Piry, S., Ray, N., Tolley, K., Duke, S., Birkun, A., Ferreira, M., Jauniaux, T., Llavona, A., Östürk, B., Östürk, A., Ridoux, V., Rogan, E., Sequeira, M., Siebert, U., Vikingson, G., Bouquegneau, J., and Michaux, J. (2007).
Rise of oceanographic barriers in continuous populations of a cetacean: the genetic structure of harbour porpoises in old world waters.
BMC Biology, 5(30).

References III

[Guillot et al., 2012] Guillot, G., Renaud, S., Ledevin, R., Michaux, J., and Claude, J. (2012).

A unifying model for the analysis of phenotypic, genetic and geographic data.

Systematic Biology, 61.

[Hannelius et al., 2008] Hannelius, U., Salmela, E., Lappalainen, T., Guillot, G., Lindgren, C., von Döbeln, U., Lahermo, P., and Kere, J. (2008).

Population substructure in Finland and Sweden revealed by a small number of unlinked autosomal SNPs.

BMC Genetics, 9(54).

[Joost et al., 2007] Joost, S., Bonin, A., Bruford, M., Després, L., Conord, C., Erhardt, G., and Taberlet, P. (2007).

A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation.

Molecular Ecology, 16:3955–3969.

References IV

[Joost et al., 2008] Joost, S., Kalbermatten, M., and Bonin, A. (2008).

Spatial analysis method (SAM): a software tool combining molecular and environmental data to identify candidate loci for selection.

Molecular Ecology Resources, 8:957–960.

[Lindgren et al., 2011] Lindgren, F., Rue, H., and Lindström, E. (2011).

An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach.

Journal of the Royal Statistical Society, series B, 73(4):423–498.

[Macholán et al., 2011] Macholán, M., Baird, S. J. E., Dufková, P., Munclinger, P., Bímová, B. V., and Piálek, J. (2011).

Assessing multilocus introgression patterns: a case study on the mouse X chromosome in central Europe.

Evolution.

DOI: 10.1111/j.1558-5646.2011.01228.

References V

[Nicholson et al., 2002] Nicholson, G., Smith, A., Jónsson, F., Gústafsson, Ó., Stefánsson, K., and Donnelly, P. (2002).

Assessing population differentiation and isolation from single-nucleotide polymorphism data.

Journal of the Royal Statistical Society, series B, 64(4):695–715.

[Pritchard et al., 2000] Pritchard, J., Stephens, M., and Donnelly, P. (2000).

Inference of population structure using multilocus genotype data.

Genetics, 155:945–959.

[Richardson and Clifford, 1991] Richardson, S. and Clifford, P. (1991).

Testing association between spatial processes.

Lecture Notes-Monograph Series, 20:295–308.

References VI

[Rosenberg et al., 2002] Rosenberg, N., Pritchard, J., Weber, J., Cann, H., Kidd, K., Zhivotovsky, L., and Feldman, M. (2002).

Genetic structure of human populations.

Science, 298:2981–2985.

[Rue et al., 2009] Rue, H., Martino, S., and Chopin, N. (2009).

Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations.

Journal of the Royal Statistical Society, series B, 71(2):1–35.

[Safner et al., 2011] Safner, T., Miller, M., McRae, B., Fortin, M., and Manel, S. (2011).

Comparison of Bayesian clustering and edge detection methods for inferring boundaries in landscape genetics.

International Journal of Molecular Sciences, 12:865–889.

References VII

[Schlather et al., 2004] Schlather, M., Ribeiro, P., and Diggle, P. (2004).

Detecting dependence between marks and locations of marked point processes.

Journal of the Royal Statistical Society, series B, 66:79–93.

[Smouse et al., 1986] Smouse, P., Long, J., and Sokal, R. (1986).

Multiple regression and correlation extensions of the Mantel test of matrix correspondence.

Systematic Zoology, 35(4):627–632.